

Telecom Churn Analysis

Pranjal Chaudhary

Data Science trainee (AlmaBetter)

Abstract:

The companies in the telecom industry face challenges when it comes to their customer churn rates. With this data analysis, we can try and find the reasons and solutions to this problem and understand the user behavior.

Problem Statement:

Orange S.A., formerly France Télécom S.A., is a French multinational telecommunications corporation. Orange Telecom's Churn Dataset consists of cleaned customer activity data (features) and a churn label specifying whether a customer canceled the subscription.

Explore and analyze the data to discover critical factors responsible for customer churn and develop ways/recommendations to ensure customer retention.

Introduction:

“Churn” concerning the Telecom industry is defined as the number/percentage of subscribers moving from a specific service or a service provider to another.

Research indicates that churn can be caused by low customer satisfaction, poor customer service, or lack of any special services that can provide monetary benefits to the user.

High churn rates (more than 10%) year on year can lead to a considerable loss in revenue and a decline in the company's reputation, leading to further repercussions.

Dataset Features:

The dataset provided consists of 20 features mentioned below.

- **State**: The state from which the customer belongs.
- **Account length**: The duration for which the customer has been with the network.
- **Area code**: The area code of the customer's neighborhood.
- **International plan**: Whether the customer has opted for the international plan or not.
- **Voice mail plan**: Whether the customer has opted for the voice mail plan or not.
- **Number vmail messages**: Number of voice mail messages the customer has received.
- **Total day minutes**: The number of minutes the customer spends on calls during the daytime.
- **Total day calls**: The number of calls the customer makes during the daytime.
- **Total day charge**: How much the customer is charged during the daytime for the calls.
- **Total eve minutes**: The total number of minutes the customer spends on calls during the evening.
- **Total eve calls**: The number of calls the customer make during evening.
- **Total day charge**: How much the customer is charged in total during the evening for the calls
- **Total night minutes**: The total number of minutes the customer spend on calls during night.
- **Total night calls**: The number of calls the customer make during the night.
- **Total night charge**: How much the customer is charged in total during night for the calls.
- **Total intl minutes**: The total number of minutes the customer spends on international calls.
- **Total intl calls**: The number of international calls the customer makes.

Total intl charge: How much the customer is charged in total for his international calls.

- **Customer service calls:** The number of calls the customer made to the customer service
- **Churn:** Whether the customer has been retained (False) or has left the network (True).

Libraries used:

- We used Pandas' library for data manipulation and presentation. • Matplotlib was used for visualizations.
- Seaborn library was used for producing visualization as well. • Google.colab library was used to import the data and save plots locally.

Steps involved:

- First, the data was loaded into notebooks using google. colab library.
- Then we performed an introductory analysis of the data by exploring the different features and their descriptions.
- The next step was data cleaning, in which we checked the dataset for the presence of any duplicate or null values.
We renamed the columns to further help in reducing errors while writing code.
- Then we added new features such as the total minutes spent on the network and charge per minute for further analysis and insights.
- Next, the analysis was performed, which involved analyzing the demographic and exploring any correlations in the data that can help understand customers' churn rate behavior.

Conclusion:

After the EDA of the dataset, we can see that the company does face an issue of high churn rates, with nearly 15 % of customers have left the network. Some correlation was observed between customer service calls and churn, indicating a need to focus on improving problem resolution speed. Another area that can improve is how much the customers are being charged for their daytime calls, as nearly 6% more users who spend more than the average amount on daytime calls have left the network compared to users who don't. Also, we observed that customers were being charged the same amount for their international calls per minute irrespective of whether they had opted for the international plan, which led to nearly 40% of customers with international plans leaving the network.

On the other side, the customer churn percentage was nearly half for customers who opted for the voice mail plan compared to customers who didn't.

Proposed solutions can include:

- 1) The customer service executives should be trained to provide excellent service quality and should be able to resolve problems quickly (preferably in the first call.
- 2) New attractive voice mail plans should be introduced, leading to a higher retention rate, as seen in the data.
- 3) There is a margin to decrease charge per minute for day caller, as it makes the service more attractive for customers
- 4) The international plan needs to be remodeled completely to be competitive in the market and provide some monetary benefits to existing and potential customers.

Resources:

- 1)Pandas Documentation
- 2)Matplotlib documentation
- 3)Wikipedia